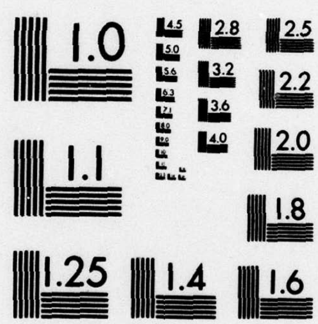
AD-A073 118 PRINCETON UNIV N J DEPT OF STATISTICS F/G 12/1
PREDICTING UNOBSERVABLE VALUES AND ESTIMATING MISSING ONES. A C--ETC(U)
JUN 79 A HOUTMAN N00014-79-C-0322
UNCLASSIFIED TR-148-SER-2-PT-1 NL

1 OF 1
AD
A073118

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A073118

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Tech. Report No. 148, Ser.2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) PREDICTING UNOBSERVABLE VALUES AND ESTIMATING MISSING ONES. A COORDINATE-FREE APPROACH. Part 1.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Anne Houtman		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-79-C-0322 New
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Princeton University Princeton, N.J. 08544		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (12) 19p.
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 436) Arlington, Virginia 22217		12. REPORT DATE (11) June 1979
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (14) TR-148-SER-2-PT-1		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE Unclassified

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A coordinate-free approach to linear prediction is developed. The results extend to those obtained by Watson (1972) and they are applied here to the estimation of random effects in mixed linear models. The treatment is self-contained and gives a study of when the BLU and LS predictors are identical. In Part II the methods will be applied to the closely related topic of estimating missing values in the analysis variance.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406 873

JLB

LEVEL IV

6

PREDICTING UNOBSERVABLE VALUES AND ESTIMATING MISSING ONES

A COORDINATE-FREE APPROACH

PART 1.

by

Anne Houtman
Princeton University

Technical Report No. 148, Series 2
Department of Statistics
Princeton University
June 1979

Research supported in part by a contract with the Office of Naval Research, No. N00014-79-C-0322, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

SUMMARY

A coordinate-free approach to linear prediction is developed. The results extend those obtained by Watson (1972) and they are applied here to the estimation of random effects in mixed linear models. The treatment is self-contained and gives a study of when the BLU and LS predictors are identical. In part II the methods will be applied to the closely related topic of estimating missing values in the analysis variance.

§1 Introduction

Problems in linear models or in the analysis of variance gain in clarity when *they are perceived as simple extensions of the old Pythagorean theorem.* The coordinate-free approach to univariate linear models provides an immediate means of access to such a geometrical picture, and allows a unified approach to problems which are usually thought of as different or unrelated such as the estimation of missing or mixed-up observations, or the fitting of models when extra data are present in the analysis of variance. Indeed these problems formed the basis of Kruskal's (1961) early advocacy of the coordinate-free viewpoint, and the purpose of these papers is to show how problems involving the prediction of unobserved or unobservable random variables fit into the same picture.

The main results of this first paper are shown to include ones concerning linear prediction within a classical linear model framework, due to Watson (1972), as well as others due to Henderson (1963), Searle (1974) and Harville (1976) relating to the "estimation" of random effects in mixed linear models. With this background, the estimation of missing or mixed-up values in designed experiments naturally suggests itself, and in the second paper we prove that the solution to this problem is identical to the best predictor of the unobserved values. This result extends and proves an observation of Fairfield-Smith (1957).

We start in section 2 with some basic definitions and properties of linear models with a possibly singular covariance matrix. Our main sources here are Kruskal (1961, 1968), Drygas (1970), Eaton (1972) and Rao (1974). Section 3 gives a derivation of the best linear unbiased predictor (BLUP) for a vector of new values. It generalizes Watson's (1972) result in that the treatment span (model subspace) is allowed to intersect with the null space of the covariance *operator and its main difference with Harville's (1974) approach is our avoidance* of the Zyskind-Martin class of g-inverses. In section 4 we obtain conditions under which the BLUP is identical to the LSP, and the paper concludes by making the connection with the classical approach to prediction and with the treatment of mixed linear models.

§2 A coordinate-free approach to univariate linear models

Let $\mathcal{D}$ be a finite dimensional vector space (the data space) of dimension n, and $\mathcal{T}$ a linear manifold of $\mathcal{D}$. The space $\mathcal{D}$ is endowed with the inner product

$$\langle x,y\rangle = x^*y$$

where x^* denotes the transpose of the column vector x. Other inner products will be defined on $\mathcal{D}$ later on. In particular, since the covariance operator will be allowed to be singular, it will be necessary to introduce a semi-inner product, that is, a nonnegative inner product.

A vector of n scalar observations is a realization of a $\mathcal{D}$-valued random vector and we make a slight departure from conventional notations in that instead of using capital letters for the observed values, lower case letters will be used for all vectors and capital letters for linear operators (or their matrix representations). Linear spaces will be denoted by script letters with the corresponding capital letters for the orthogonal projections onto them with respect to $\langle .,.\rangle$ defined above.

The expectation $\mathbb{E}y$ of y is the unique vector τ of $\mathcal{D}$ such that for all $a \in \mathcal{D}$

$$\mathbb{E}\langle a,y\rangle = \langle a,\tau\rangle \ .$$

This definition does not depend on the choice of a particular inner product on $\mathcal{D}$. The vector τ will be assumed to lie in the subspace $\mathcal{T}$ of $\mathcal{D}$.

The covariance $\operatorname{cov}(y)$ of y is the unique linear operator V on $\mathcal{D}$ such that for all $a,b \in \mathcal{D}$

$$\operatorname{cov}[\langle a,y\rangle,\langle b,y\rangle] = \langle a,Vb\rangle \ .$$

The operator V is nonnegative definite and symmetrical, hence $\langle a,Vb\rangle$ is a semi-inner product on $\mathcal{D}$, we will denote it by $\langle a,b\rangle_V = \langle a,Vb\rangle$. We have $\operatorname{var}[\langle a,y\rangle] = \langle a,Va\rangle = \langle a,a\rangle_V = \|a\|_V^2$. Clearly the definition of V does depend on the inner product used.

For any subspaces A, $\mathcal{B}$ of $\mathcal{D}$ and linear operator C, we write $A^{\perp} = \{z \in \mathcal{D} : \langle z,a\rangle = 0 \text{ for all } a \in A\}$, $CA = \{Ca : a \in A\}$, $A + \mathcal{B} = \{a + b : a \in A, b \in \mathcal{B}\}$ and $A + \mathcal{B}$ is written $A \oplus \mathcal{B}$ when $A \cap \mathcal{B} = \{0\}$. Finally we denote the kernel (or null space) of C by $\mathcal{K}(C) = \{z \in \mathcal{D} : Cz = 0\}$ and the range of C by $\mathcal{R}(C) = \{Cz : z \in \mathcal{D}\}$. We now prove the following results due to Rao (1974).

Lemma 2.1 Let y be a random vector with $\mathbb{E}y \in \mathcal{T}$ and $\mathrm{cov}(y) = V$. Then y belongs to $\mathcal{T} + \mathcal{R}(V)$ with probability one.

Proof: We first note that $\mathcal{T} + \mathcal{R}(V) = (\mathcal{T}^{\perp} \cap \mathcal{K}(V))^{\perp}$. Then, for all $z \in \mathcal{T}^{\perp} \cap \mathcal{K}(V)$ we have

$$\mathbb{E}\langle z,y\rangle = \langle z,\tau\rangle = 0 \ ; \quad \mathrm{var}\langle z,y\rangle = \langle z,Vz\rangle = 0 \ .$$

Hence for all $z \in \mathcal{T}^{\perp} \cap \mathcal{K}(V)$, $\langle z,y\rangle = 0$ with probability one. This shows that $y \in (\mathcal{T}^{\perp} \cap \mathcal{K}(V))^{\perp}$ with probability one or, equivalently, that $y \in \mathcal{T} + \mathcal{R}(V)$ with probability one.□

Lemma 2.2 $\mathcal{T} + \mathcal{R}(V) = \mathcal{T} \oplus V\mathcal{T}^{\perp}$.

Proof: We first show that $\mathcal{T} \cap V\mathcal{T}^{\perp} = \{0\}$. If $V\bar{T}b \in \mathcal{T}$, where $\bar{T} = I - T$, then $\|V^{1/2}\bar{T}b\|^2 = \langle V^{1/2}\bar{T}b, V^{1/2}\bar{T}b\rangle = \langle b, \bar{T}V\bar{T}b\rangle = 0$ whence $V\bar{T}b = 0$. Then we have $\mathcal{T} \oplus \mathcal{R}(\bar{T}V) = \mathcal{T} \oplus \bar{T}\mathcal{R}(V) = \mathcal{T} + \mathcal{R}(V) \supseteq \mathcal{T} \oplus V\mathcal{R}(\bar{T}) = \mathcal{T} \oplus \mathcal{R}(V\bar{T})$. But since $\mathrm{rank}(V\bar{T}) = \mathrm{rank}(\bar{T}V)$ we must have equality.□

When $\mathcal{T} \oplus V\mathcal{T}^{\perp}$ is a proper subspace of the Euclidean space of dimension n, i.e. when V is singular, we take the data space $\mathcal{D}$ to be that subspace rather than the whole space. Now every direct sum decomposition of a vector space defines projections <u>onto</u> one of the components <u>along</u> the other and in our case we write these as $P_{\mathcal{T}|V\mathcal{T}^{\perp}}\, x = x_1$ and $P_{V\mathcal{T}^{\perp}|\mathcal{T}}\, x = x_2$, where

$$x = x_1 + x_2\,,\quad x_1 \in \mathcal{T},\ x_2 \in V\mathcal{T}^{\perp}$$

is the unique decomposition of $x \in \mathcal{D} = \mathcal{T} \oplus V\mathcal{T}^{\perp}$. The range space and kernel of $P_{\mathcal{T}|V\mathcal{T}^{\perp}}$ and $P_{V\mathcal{T}^{\perp}|\mathcal{T}}$ are clearly $\mathcal{T}$ and $V\mathcal{T}^{\perp}$, and vice versa, respectively, and on $\mathcal{D}$ we have

$$P_{\mathcal{T}|V\mathcal{T}^{\perp}} + P_{V\mathcal{T}^{\perp}|\mathcal{T}} = I\,. \tag{1}$$

These projections are directly related to the projections $P_{\mathcal{T},V}$ of $\mathcal{D}$ onto $\mathcal{T}$ defined relative to the semi-inner product $\langle .,. \rangle_V$ by (i) $R(P_{\mathcal{T},V}) \subset \mathcal{T}$, and (ii) for all $y \in \mathcal{D}$ and $t \in \mathcal{T}$, $\|y - P_{\mathcal{T},V}y\|^2 \le \|v - t\|_V^2$; equivalently, (ii)' $TVP_{\mathcal{T},V} = TV$. The following relation is given by Rao (1974)

$$P_{\mathcal{T}|V\mathcal{T}^{\perp}} = I - P^*_{\mathcal{T}^{\perp},V} \quad \text{on } \mathcal{D}\,. \tag{2}$$

We now derive the best linear unbiased estimator (BLUE) of a linear functional $\langle t,\tau \rangle$, $\tau \in \mathcal{T}$, where, without loss of generality, we can restrict ourselves to coefficients $t \in \mathcal{T}$, since for all $z \in \mathcal{D}$, $\langle z,\tau \rangle = \langle Tz,\tau \rangle$. Any linear unbiased estimator of $\langle t,\tau \rangle$ is of the form $\langle z,y \rangle$ where $z \in \mathcal{D}$ is such that $Tz = t$, that is, $z = t - u$ for some $u \in \mathcal{T}^{\perp}$. Such an estimator is the BLUE of $\langle t,\tau \rangle$ if it minimizes $E[\langle z,y \rangle - \langle t,\tau \rangle]^2 = \|t - u\|_V^2$, that is, using the definition above, if $\hat{u} = P_{\mathcal{T}^{\perp},V}\,t$. Hence the BLUE of $\langle t,\tau \rangle$ is $\langle t - P_{\mathcal{T}^{\perp},V}\,t, y \rangle$ or, using (2),

$$\langle t, P_{\mathcal{T}|V\mathcal{T}^{\perp}}\,y \rangle\,.$$

By letting t range over an orthonormal basis of $\mathcal{T}$ it is easily seen that the BLUE of $Ey = \tau$ is $P_{\mathcal{T}|V\mathcal{T}^{\perp}}\,y$, and this estimator is uniquely determined on $\mathcal{D}$.

When V is not singular, in which case $\mathcal{D}$ is the entire Euclidean space, the BLUE $P_{\mathcal{J}|V\mathcal{J}^\perp}y$ is identical to $P_{\mathcal{J},V^{-1}}y$, the orthogonal projection of the data onto $\mathcal{J}$ w.r.t. $\langle\cdot,\cdot\rangle_{V^{-1}}$.

The following result gives a necessary and sufficient condition for the BLUE of Ey to be identical to the least squares estimator (LSE) of Ey. The proof is particularly simple.

<u>Lemma 2.3</u> The projection onto $\mathcal{J}$ along $V\mathcal{J}^\perp$ is identical to the orthogonal projection T onto $\mathcal{J}$ if and only if V leaves $\mathcal{J}$ (or equivalently $\mathcal{J}^\perp$) invariant.

Proof: Clearly if V leaves $\mathcal{J}^\perp$ invariant, then $P_{\mathcal{J}|V\mathcal{J}^\perp} = P_{\mathcal{J}|\mathcal{J}^\perp} = T$. Conversely, if $T = P_{\mathcal{J}|V\mathcal{J}^\perp}$, then $\overline{T} = P_{V\mathcal{J}^\perp|\mathcal{J}}$; this implies $\mathcal{R}(\overline{T}) = \mathcal{R}(P_{V\mathcal{J}^\perp|\mathcal{J}})$ or, equivalently, $\mathcal{J}^\perp = V\mathcal{J}^\perp$. □

Problems of best linear unbiased estimation with singular covariance matrices were approached differently by Philoche (1971) who first defined an inner product restricted on $\mathcal{R}(V)$, $\langle\langle Vx,Vy\rangle\rangle_V = \langle Vx,y\rangle$ which is identical to $\langle\cdot,\cdot\rangle_{V^{-1}}$ on $\mathcal{R}^n$ when V is not singular. Philoche then defines the class of Gauss-Markov operators, that is, linear operators that are the identity on $\mathcal{J}$ and whose restriction to $\mathcal{R}(V)$ coincides with the orthogonal projection from $\mathcal{R}(V)$ onto $\mathcal{R}(V) \cap \mathcal{J}$ w.r.t. $\langle\langle\cdot,\cdot\rangle\rangle_V$. When the discussion is restricted to the space $\mathcal{J} + \mathcal{R}(V)$ to which the data belongs with probability one, we find out that all the operators in the class defined by Philoche coincide with the unique operator $P_{\mathcal{J}|V\mathcal{J}^\perp}$.

§3. Linear prediction

Assume that we do not observe all of y, but only a "part" y_1 of it and we want to use the observations y_1 to predict y. More precisely, suppose that $\mathcal{D}$ is the direct sum of two orthogonal subspaces

$$\mathcal{D} = \mathcal{D}_1 \oplus \mathcal{D}_2$$

reflecting the decomposition

$$y = y_1 + y_2$$

of the full data vector y into the observed part $y_1 = D_1 y$ and the unobserved part $y_2 = D_2 y$. With this notation our model becomes

$$\mathbb{E} y_i \in \mathcal{T}_i \tag{3}$$

$$\operatorname{cov}(y_i) = D_i V D_i \qquad i = 1, 2 \tag{4}$$

$$\text{and } \operatorname{cov}(y_1, y_2) = D_1 V D_2 , \tag{5}$$

where $\mathcal{T}_i = D_i \mathcal{T}$ are the model subspaces for the observed and unobserved data respectively.

The problem is to find the best linear unbiased predictor (BLUP) of y_2 (or equivalently of y) based on the realized values of y_1, i.e. to find A such that $\mathbb{E}\|Ay_1 - y_2\|^2$ is minimum subject to $\mathbb{E}Ay_1 = \mathbb{E}y_2$. Our approach to find the BLUP of y_2 is as follows: (i) we form an estimate $\hat{\tau}_1$ of the mean $\tau_1 = D_1\tau$ of y_1; (ii) then we use this to get an estimate $\hat{\tau}_2$ of the mean $\tau_2 = D_2\tau$ of y_2; (iii) finally we use the deviation $y_1 - \hat{\tau}_1$ of y_1 from its estimated mean to give an indication of $y_2 - \tau_2$. The BLUP of y_2 will then be of the form $\hat{\tau}_2 + C(y_1 - \hat{\tau}_1)$.

We will need the following two lemmas. The first lemma is a general result on symmetric nonnegative operators for which a proof can be found in Eaton (1972, proposition 3.31). The second lemma is an extension of proposition 4.3 in Eaton (1972) to the case of a general nonnegative covariance operator.

Lemma 3.1 $\mathcal{R}(D_1VD_2) \subset \mathcal{R}(D_1VD_1)$ and symmetrically $\mathcal{R}(D_2VD_1) \subset \mathcal{R}(D_2VD_2)$.

Lemma 3.2 If z is a random vector with $\mathbb{E}z \in A$ and nonnegative definite $\mathrm{cov}(z) = W$, and if F is a linear operator from A to a linear space $\mathcal{B}$, then the BLUE of $F\mathbb{E}z$ is $F\widehat{\mathbb{E}z}$ where $\widehat{\mathbb{E}z}$ is the BLUE of $\mathbb{E}z$.

We now make the following important assumption:

$$\dim \mathcal{J}_1 = \dim \mathcal{J} \tag{6}$$

A consequence of (6) is that for every $\tau_1 \in \mathcal{J}_1$, there is a unique $\tau = M\tau_1 \in \mathcal{J}$ such that $D_1\tau = \tau_1$. This correspondence is clearly linear so we can define an operator $M : \mathcal{D}_1 \to \mathcal{D}$ such that $Md = 0$ if $d \in \mathcal{D}_1 \ominus \mathcal{J}_1$ and $D_1M\tau_1 = \tau_1$ for all $\tau_1 \in \mathcal{J}_1$. Now by lemma 2.1 y_1 is restricted to the linear space $\mathcal{J}_1 \oplus (D_1VD_1)\mathcal{J}_1^{\perp}$ with probability one and so the unique BLUE of $\mathbb{E}y_1 = \tau_1$ is $\hat{\tau}_1 = P_{\mathcal{J}_1|(D_1VD_1)\mathcal{J}_1^{\perp}} y_1$.

It follows from lemma 3.2 that $M\hat{\tau}_1$ is the BLUE of $\mathbb{E}y = \tau$ using y_1 , and that $D_2M\hat{\tau}_1$ is the BLUE $\hat{\tau}_2$ of $\mathbb{E}y_2 = D_2\tau$ based on the observed data y_1 .

With these preliminaries we can now give our theorem, the assumptions and notations being as in (3), (4), (5) and (6).

Theorem 3.1 The BLUP of y_2 based on y_1 has the form

$$\hat{y}_2 = \hat{\tau}_2 + \left(D_2VD_1\right)\left(\sum_{i=1}^{k} f_i^{-1}F_i\right)\left(y_1 - \hat{\tau}_1\right) \tag{7}$$

where $\sum_{i=1}^{k} f_i^{-1}F_i$ is a symmetric g-inverse of $D_1VD_1 = \sum_{i=0}^{k} f_iF_i$, $\hat{\tau}_1$ is the BLUE of $\mathbb{E}y_1$ and $\hat{\tau}_2 = D_2M\hat{\tau}_1$ is the BLUE of $\mathbb{E}y_2$ based on y_1 .

Proof: For any predictor of the form Ay_1 the unbiasedness condition $\mathbb{E}Ay_1 = \mathbb{E}y_2$ is equivalent to $A\tau_1 = D_2M\tau_1$ for all $\tau_1 \in \mathcal{J}_1$, i.e. $AT_1 = D_2MT_1$. This implies the following decomposition, where $\overline{T}_1 = I - T_1$:

$$Ay_1 = AT_1y_1 + A\overline{T}_1y_1 ,$$

or

$$Ay_1 = D_2MT_1y_1 + B\overline{T}_1y_1 . \tag{8}$$

Further, it can be assumed without loss of generality that

$$\mathcal{R}(B^*) \subset \mathcal{R}(\overline{T}_1(D_1VD_1)\overline{T}_1) ; \tag{9}$$

indeed $\mathbb{E}\overline{T}_1y_1 = 0$ and $\operatorname{cov}(\overline{T}_1y_1) = \overline{T}_1(D_1VD_1)\overline{T}_1$, hence by lemma 2.1 $\overline{T}_1y_1 \in \mathcal{R}(\overline{T}_1(D_1VD_1)\overline{T}_1)$.

Now the mean square error $\mathbb{E}\|Ay_1 - y_2\|^2$ can be expanded as

$$\mathbb{E}\|(D_2MT_1D_1 + B\overline{T}_1D_1 - D_2)y\|^2 =$$

$$\operatorname{tr}[D_2MT_1(D_1VD_1)T_1M^*D_2] + 2\operatorname{tr}[D_2MT_1(D_1VD_1)\overline{T}_1B^*] + \operatorname{tr}[B\overline{T}_1(D_1VD_1)\overline{T}_1B^*]$$
$$+ \operatorname{tr}[D_2VD_2] - 2\operatorname{tr}[D_2MT_1D_1VD_2] - 2\operatorname{tr}[B\overline{T}_1D_1VD_2] .$$

A standard argument using derivatives or completing the square shows that B corresponding to the minimum mean square error satisfies

$$B\overline{T}_1(D_1VD_1)\overline{T}_1 = D_2VD_1\overline{T}_1 - D_2MT_1(D_1VD_1)\overline{T}_1$$

and hence, using (9),

$$B = (D_2VD_1\overline{T}_1 - D_2MT_1(D_1VD_1)\overline{T}_1)\sum_{i=1}^{p} e_j^{-1}E_j$$

where $\sum_{j=1}^{p} e_j^{-1}E_j$ is the symmetric g-inverse of $\overline{T}_1(D_1VD_1)\overline{T}_1 = \sum_{j=0}^{p} e_jE_j$ in spectral form, with $e_0 = 0$.

Substituting B into (8) gives, after regrouping terms,

$$\hat{y}_2 = D_2MT_1\left\{I - \left(D_1VD_1\right)\overline{T}_1\left(\sum_{j=1}^{p} e_j^{-1}E_j\right)\overline{T}_1\right\}y_1 + \left(D_2VD_1\right)\overline{T}_1\left(\sum_{j=1}^{p} e_j^{-1}E_j\right)\overline{T}_1 y_1 \ . \qquad (10)$$

Letting D_1VD_1 have spectral decomposition

$$D_1VD_1 = \sum_{i=0}^{k} f_i F_i \ , \ f_0 = 0 \ ,$$

then by lemma 3.1 $\sum_{i=1}^{k} F_i = \left(\sum_{i=1}^{k} f_i^{-1}F_i\right)D_1VD_1$ is the identity on $\mathcal{R}(D_1VD_2)$, and so

$$D_2VD_1 = D_2VD_1\left(\sum_{i=1}^{k} f_i^{-1}F_i\right)D_1VD_1 \ . \qquad (11)$$

By checking that $\overline{T}_1\left(\sum_{j=1}^{p} e_j^{-1}E_j\right)\overline{T}_1(D_1VD_1)$ satisfies conditions (i) and (ii)' (section 2) we see that it coincides with $P_{\mathcal{J}_1^{\perp},D_1VD_1}$, and so by using (2) we have

$$\left(D_1VD_1\right)\overline{T}_1\left(\sum_{j=1}^{p} e_j^{-1}E_j\right)\overline{T}_1 = I - P_{\mathcal{J}_1|(D_1VD_1)\mathcal{J}_1^{\perp}} \ . \qquad (12)$$

Hence, by substituting (11) and (12) into (10), we have

$$\hat{y}_2 = D_2MT_1P_{\mathcal{J}_1|(D_1VD_1)\mathcal{J}_1^{\perp}}y_1 + \left(D_2VD_1\right)\left(\Sigma f_i^{-1}F_i\right)\left(y_1 - P_{\mathcal{J}_1|(D_1VD_1)\mathcal{J}_1^{\perp}}y_1\right)$$

and, since T_1 is the identity on $\mathcal{J}_1$ and $P_{\mathcal{J}_1|(D_1VD_1)\mathcal{J}_1^{\perp}}y_1 = \hat{\tau}_1$ is the BLUE of Ey_1 , the proof is complete.□

§4. When are the BLUP and the LSP identical?

It is easy to verify that the least squares predictor (LSP) of y_2 based on y_1 is

$$\tilde{y}_2 = D_2 M T_1 y_1$$

that is, the LSP of y_2 is identical to the LS estimator of $\mathbb{E}y_2$ based on y_1. This expression is much simpler to evaluate than the expression for the BLUP. Moreover the LSP does not involve the usually unknown covariance operator V. It is thus of interest to look for conditions under which the BLUP can be replaced by the simpler expression for the LSP. The following result generalizes a theorem proved by Watson (1972); the framework at the beginning of section 3 continues to apply and $\mathcal{T}_1$ is assumed to be a fixed subspace of $\mathcal{D}_1$.

Theorem 4.1 For every $\mathcal{T} \subset \mathcal{D}$ such that $D_1\mathcal{T} = \mathcal{T}_1$ and $\dim \mathcal{T} = \dim \mathcal{T}_1$, the LSP of $y_2 \in \mathcal{D}_2$ coincides with the BLUP of y_2 if and only if

$$\text{(i)} \qquad (D_1 V D_1)\mathcal{T}_1 \subset \mathcal{T}_1$$

$$\text{(ii)} \qquad \mathcal{R}(D_1 V D_2) \subset \mathcal{T}_1 .$$

Proof: First assume that (i) and (ii) hold. By lemma 2.3 $P_{\mathcal{T}_1 | (D_1 V D_1)\mathcal{T}_1^{\perp}} y_1 = T_1 y_1$ and hence $\hat{\tau}_1 = \tilde{y}_2$, and by using the expression of $\hat{y}_2$ given in (10) it is easy to see that the second term is zero if (ii) holds. Conversely, assume that for any $\mathcal{T}$ such that $\dim \mathcal{T} = \dim \mathcal{T}_1$ we have

$$D_2 M T_1 y_1 = D_2 M \hat{\tau}_1 + \left(D_2 V D_1\right)\left(\sum_{i=1}^{k} f_i^{-1} F_i\right)\left(y_1 - \hat{\tau}_1\right).$$

Since this identity must hold in particular for $\mathcal{T} = \mathcal{T}_1$, that is for $\mathcal{T}_2 = \{0\}$, we must have (a) $D_2 M T_1 y_1 = D_2 M \hat{\tau}_1$ and (b) $(D_2 V D_1)(\Sigma f_i^{-1} F_i)(y_1 - \hat{\tau}_1) = 0$.

Relation (a) must hold in particular for $\mathcal{J}_2$ such that $\dim \mathcal{J}_2 = \dim \mathcal{J}$ and so $T_1 y_1 = \hat{\tau}_1$ since M is bijective. This implies that $(D_1 V D_1)\mathcal{J}_1 \subset \mathcal{J}_1$ by lemma 4.1. Now using the expression of the second term in (10) and the fact that $(D_1 V D_1)\mathcal{J}_1^{\perp} \subset \mathcal{J}_1^{\perp}$ we can conclude that $\mathcal{J}_1^{\perp} \subset \mathcal{N}(D_2 V D_1)$ or equivalently that $\mathcal{R}(D_2 V D_1) \subset \mathcal{J}_1$. This completes the proof.□

§5. Applications

To show how our results relate to classical results let us specify a basis in $\mathcal{D}$ such that vectors are arrays of coordinates in that basis, and linear operators are given a matrix representation.

Watson (1972) considered a linear model

$$y_1 = X_1\beta + \varepsilon_1$$

where y_1 and ε_1 are $n \times 1$ random vectors, X_1 is a known and possibly singular $n \times k$ design matrix, β is a $k \times 1$ vector of unknown parameters, $E\varepsilon_1 = 0$ and $\operatorname{var} \varepsilon_1 = \Gamma$, a nonnegative symmetric matrix. The problem was to find the BLUP of

$$y_2 = x_2^*\beta + \varepsilon_2$$

where the column vector x_2 belongs to the column space $\mathcal{C}(X_1^*)$ of X_1^* (the row space of X_1), $\operatorname{var} \varepsilon_2 = \sigma^2 > 0$ and $\operatorname{cov}(y_1, y_2) = \gamma$. These equations can be expressed in the following way:

$$\begin{bmatrix} y_1 \\ \hline 0 \end{bmatrix} = \begin{bmatrix} X_1 \\ \hline 0 \; . \; . \; . \; 0 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \hline 0 \end{bmatrix} ; \quad \begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \\ \hline y_2 \end{bmatrix} = \begin{bmatrix} 0 \; 0 \; . \; . \; 0 \\ . \qquad . \\ . \qquad . \\ . \qquad . \\ 0 \; . \; . \; . \; 0 \\ \hline x_2^* \end{bmatrix} \beta + \begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \\ \hline \varepsilon_2 \end{bmatrix}$$

Letting $y^* = (y_1^* \;\; y_2^*)$, $X^* = (X_1^* \;\; x_2^*)$ and $\varepsilon^* = (\varepsilon_1^* \;\; \varepsilon_2^*)$ we can write these as

$$y = X\beta + \varepsilon$$

where in the notation of section 3, $Ey \in \mathcal{C}(X) = \mathcal{J}$, and

$$\mathcal{C}\begin{bmatrix} X_1 \\ \hline 0 \ . \ . \ . \ 0 \end{bmatrix} = \mathcal{J}_1 \ , \ \mathcal{C}\begin{bmatrix} 0 \ . \ . \ . \ 0 \\ . \\ . \\ 0 \ . \ . \ . \ 0 \\ \hline x_2^* \end{bmatrix} = \mathcal{J}_2 \ , \ \left[\begin{array}{c|c} \Gamma & \gamma \\ \hline \gamma' & \sigma^2 \end{array}\right] = V \ , \ \left[\begin{array}{c|c} \Gamma & 0 \\ & . \\ & . \\ & . \\ & 0 \\ \hline 0 \ . \ . \ . \ 0 & 0 \end{array}\right] = D_1 V D_1 \ ,$$

$$\left[\begin{array}{c|c} 0 \ . \ . \ . \ 0 & 0 \\ . & . \\ . & . \\ . & . \\ 0 \ . \ . \ . \ 0 & 0 \\ \hline 0 \ . \ . \ . \ 0 & \sigma^2 \end{array}\right] = D_2 V D_2 \ \text{ and } \ \left[\begin{array}{c|c} 0 \ . \ . \ . \ 0 & \\ . \qquad . & \\ . \qquad . & \\ . \qquad . & \gamma \\ 0 \ . \ . \ . \ 0 & \\ \hline 0 \ . \ . \ . \ 0 & 0 \end{array}\right] = D_1 V D_2 \ .$$

The condition $x_2 \in \mathcal{C}(X_1^*)$ is equivalent to the assumption made that $\dim \mathcal{J}_1 = \dim \mathcal{J}$. It is easy to see that if Γ has spectral decomposition $\sum_{i=0}^{r} g_i G_i$, then theorem 3.1 implies that

$$\hat{y}_2 = x_2^* \hat{\beta} + \gamma^* \left(\sum_{i=1}^{r} g_i^{-1} G_i \right) \left(y_1 - X_1 \hat{\beta} \right)$$

where $\hat{\beta}$ is a BLUE of β. This expression for the BLUP was obtained by Watson (1972) under the assumption that $\mathcal{C}(X_1) \subset \mathcal{C}(\Gamma)$, but it can be seen from theorem 3.1 that this hypothesis is not indispensable.

Other corollaries of theorem 3.1 include (i) the fact that the prediction of a vector of new values is equivalent to the separate predictions of each component of the vector, that is, the BLUP of an r-dimensional vector is the vector of BLUP's of each coordinate; (ii) a best predictor of the values of individual observations when only the sum (or the values of some linear combination) of these observations has been observed; (iii) an alternative derivation of results concerning estimation in mixed models. We now consider this last topic in greater detail.

Problems involving prediction in mixed models have been implicit in the literature on animal and plant breeding for many years, see especially Henderson (1963), and for more recent results Searle (1974), Henderson (1975) and Harville (1976). These models are usually formulated as

$$y = X\beta + Zu + \varepsilon$$

where y is a vector of n observations, X a known $n \times p$ matrix, Z a known $n \times q$ matrix, β a vector of p unknown parameters ("fixed effects") and u a vector of q random variables ("random effects"). It is usually assumed that $\mathbb{E}u = \mathbb{E}\varepsilon = \mathrm{cov}(u,\varepsilon) = 0$, $\mathrm{cov}(\varepsilon) = \Gamma$ ($n \times n$, n.n.d.) and $\mathrm{cov}(u) = \Lambda$ ($q \times q$, n.n.d.) The problem is then to "estimate" a linear combination

$$z = k^*\beta + m^*u \,, \; k \in \mathcal{C}(X^*) \,.$$

Clearly the mixed model can be reformulated as

$$\mathbb{E}y \in \mathcal{C}(X) \,, \; \mathrm{cov}(y) = Z\Lambda Z^* + \Gamma$$

and the "estimation" problem is simply that of predicting a random variable z whose mean and variance are related to that of y : the BLUE of $\mathbb{E}z$ can be deduced immediately from the BLUE of $\mathbb{E}y$, $\mathrm{var}\, z = m^*\Lambda m$ and $\mathrm{cov}(y,z) = Z\Lambda m$. By using the above correspondence between the classical and coordinate-free approaches, theorem 3.1 allows us to express the BLUP of z as

$$\hat{z} = \begin{pmatrix}\text{BLUE of } \mathbb{E}z \\ \text{based on} \\ \text{BLUE of } \mathbb{E}y\end{pmatrix} + \mathrm{cov}(z,y) \times \begin{pmatrix}\text{effective} \\ \text{inverse} \\ \text{of } \mathrm{cov}(y)\end{pmatrix} \times \begin{pmatrix}\text{residuals} \\ \text{of observed} \\ \text{model}\end{pmatrix}$$

i.e. $$\hat{z} = k^*\hat{\beta} + (m^*\Lambda Z^*)\left(\sum_{i=1}^{r} g_i^{-1} G_i\right)(y - X\hat{\beta})$$

where $\hat{\beta}$ is a BLUE of β and $\sum_{i=0}^{r} g_i G_i$ is the spectral form of $\mathrm{cov}(y)$.

We thus see that the estimation of random effects in mixed models is not a sui generis problem; it can be incorporated within the class of problems of prediction of an unobservable random variable from a realized value of one that can be observed.

We close this section by emphasizing three points in the preceding discussion. Firstly, we always need a condition to insure that the mean of the unobserved variable can be estimated. Secondly, as should be clear, when the observed and unobserved quantities are uncorrelated, the latter quantity is predicted by the BLUE of its mean. And finally we remark that a singular design matrix causes no problem in a geometrical approach, nor does the singularity of the covariance matrix provided discussion is restricted to the subspace in which the data falls with probability one.

Acknowledgements: I am grateful to Dr G.S. Watson for bringing the topic of prediction to my attention and to both him and Dr T.P. Speed for useful discussions on the subject.

References

Drygas, H. (1970). The coordinate-free approach to Gauss-Markov estimation. Springer-Verlag.

Eaton, M.L. (1972). Multivariate Analysis. University of Copenhagen.

Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. Ann. Stat. 4, 384-395.

Henderson, C.R. (1963). Selection index and expected genetic advance. Statistical genetics and plant breeding, NAS-NRC Publication No. 982, 141-163.

(1975). Best linear unbiased estimation and prediction under a selection model. Biometrics, 31, 423-447.

Kruskal, W. (1961). The coordinate-free approach to Gauss-Markov estimation and its application to missing and extra observations. Fourth Berkeley Symp. Math. Stat. Prob. 1, 435-451.

(1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. Ann. Math. Stat., 39, 70-75.

Philoche, J.-L. (1971). A propos du théorème de Gauss-Markov. Inst. Henri Poincaré, section B. VII, No. 4, 271-281.

Rao, C.R. (1974). Projectors, generalized inverses and the BLUE's. Ann. Stat., 3, 442-448.

Searle, S.R. (1974). Prediction, mixed models, and variance components. Reliability and Biometry (F. Proschan and R.J. Serfling, eds.),229-266. SIAM, Philadelphia.

Smith, H.F. (195[illegible] Missing plot estimates. Note 125. Biometrics 13, 115-118.

Von Neumann, J. (1950). Functional operators, vol.II: The geometry of orthogonal spaces. Princeton University Press.

Watson, G.S. (1972). Prediction and the efficiency of least squares. Biometrika, 59, 91-98.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Ann. Math. Stat., 38, 1092-1109.

Zyskind, G. and Martin, F.B. (1969). On best linear estimation and a general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure. SIAM J. Appl. Math., 17, 1190-1202.